

Community Proteogenomics:

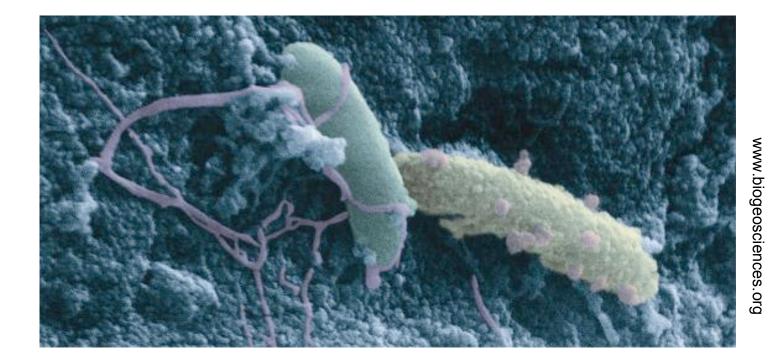
background and application to the Rifle Bioremediation project

Rifle project proteogenomics subgroup Jill Banfield, presenter

Goal:

to understand the functioning of natural microbial consortia

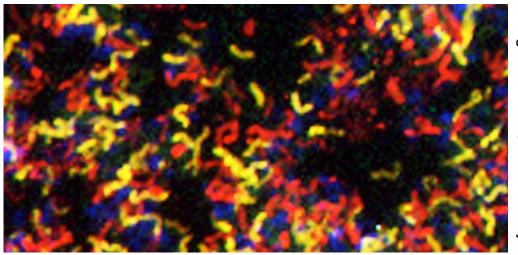
- study of coexisting organisms, not monocultures
- cultivation independent

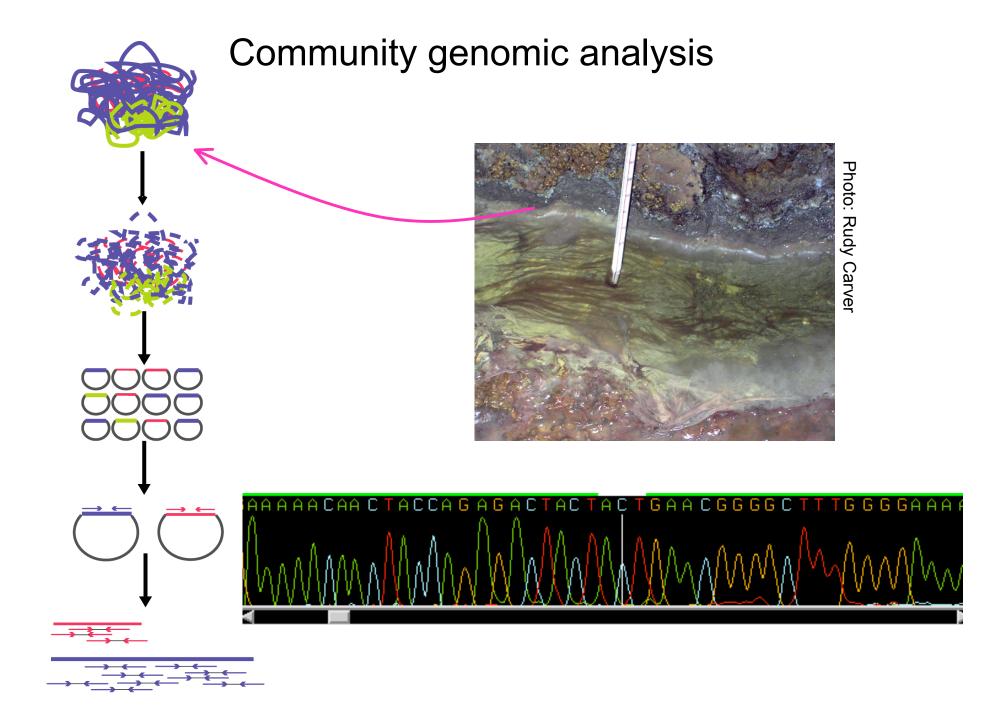


Molecular foundation:

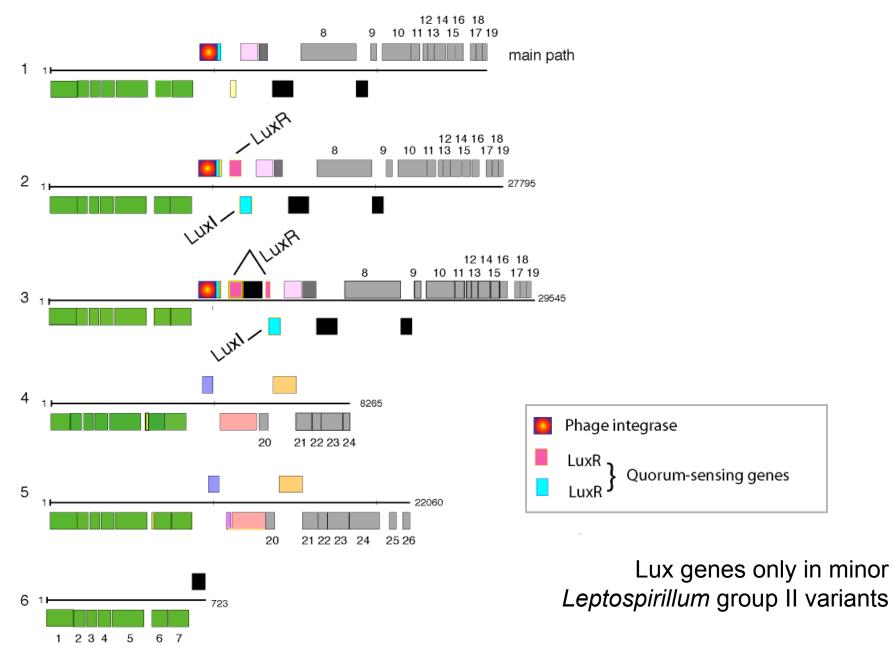
Genomes recovered from environmental samples

Metagenomics - data of data genomics or transcendential genomics Environmental genomics - the study of environmentally-derived DNA Community genomics - genomic studies of microbial communities

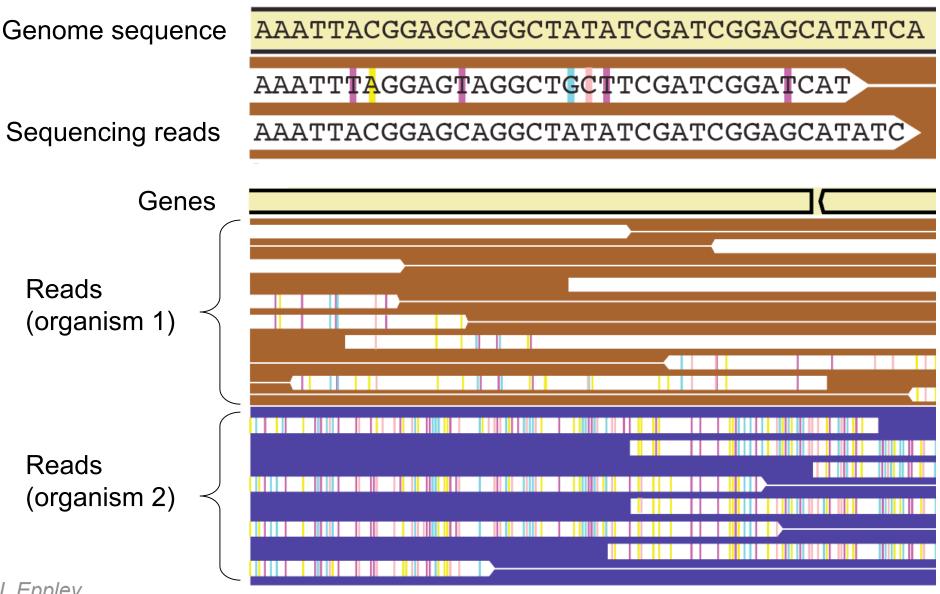




STRAIN DIFFERENCES IN GENE CONTENT

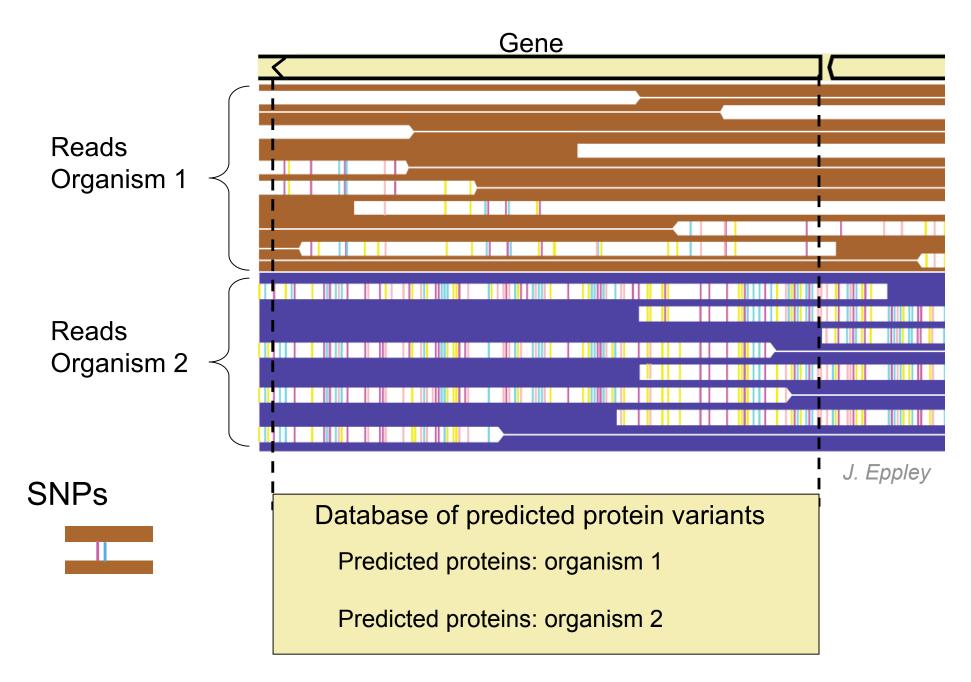


Definition of populations: variaton within vs. between clusters:



J. Eppley

Reconstruction of protein variants using "Strainer" (Eppley et al.)



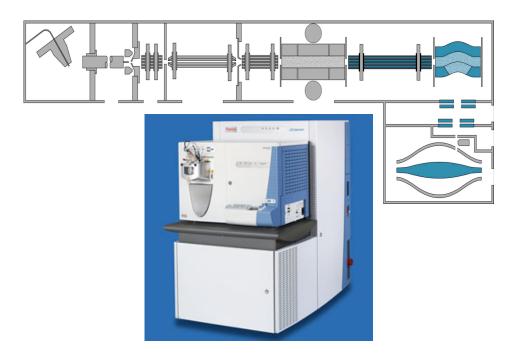
Global proteomics: Simultaneous characterization of all cell proteins

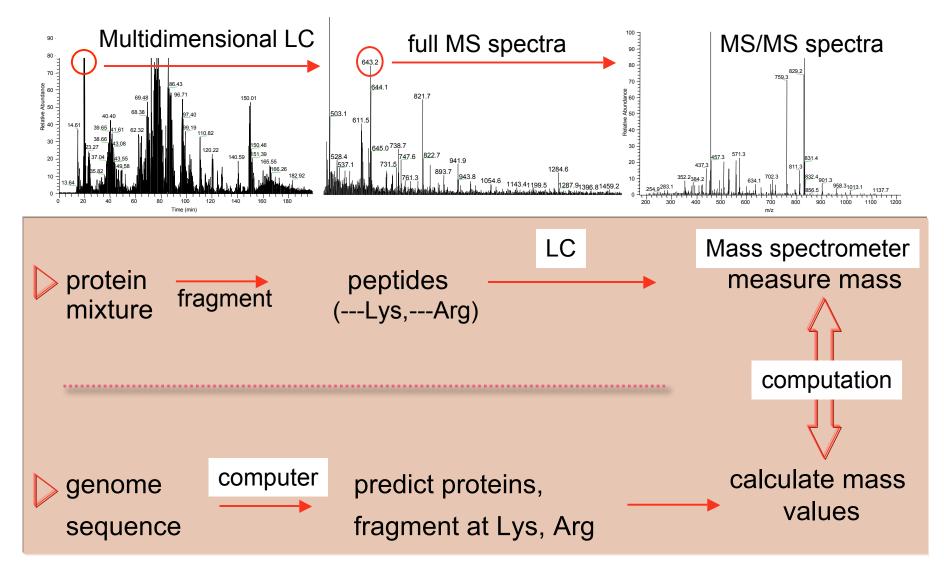
Peptide identification based on:

MS/MS fragmentation patterns of chromatographically-separated peptides; identification usually by comparison with protein fragmentation patterns predicted from genomic data

- In some cases, coupled to very high measured mass accuracy measurements of peptides in elution profiles







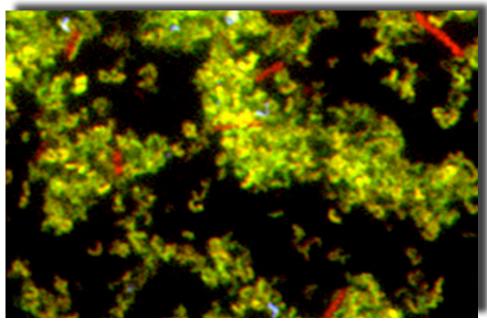
Leptospirillum group II_scaffold_14_GENE_20

2032 proteins from one biofilm

MNKWAGAVLGTVTLGLLSATAYSAELDILKPNRVPADQIAAAKAMKPPFPVTA AVIAKGKEVFNGAGTCYTCHGVGGK<mark>GDGPGAAGMDPSPR</mark>FTNHQFDQVRTAGE MVWVVSNGSPLQPAMVGFVSAGITDKQAWEAVMYERSLGCGGDMDC.....

LOW DIVERSITY COMMUNITIES AS MODEL SYSTEMS





Most biofilms dominated by:

Leptospirillum group II Leptospirillum group III

Ferroplasma types I and II A, B, C, D, E, G, H, I-plasma ARMAN 1,2,3,4 at low abundance

Blue = NEAR-COMPLETE COMPOSITE GENOMES RECONSTRUCTED OR ONGOING

> Tyson et al. *Nature* (2004) Allen et al. *PNAS* (2007) Lo et al. *Nature* (2007)

SOME SIMPLE METRICS

For the most abundant organism - single sample:

~ 52% predicted proteins detected in a single sample

For the most abundant organism - across 6 samples:

68% of all predicted proteins80% of core chromosomal proteins34% of integrated plasmid/phage proteins

Overall, proteins identified across multiple samples:

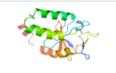
68% Leptospirillum group II
41% Leptospirillum group III
37% Gplasma
20% Ferroplasma type I
20% Ferroplasma type II

Biofilms: 2,275 - 3,989 proteins with ~1% false positive identification rate)

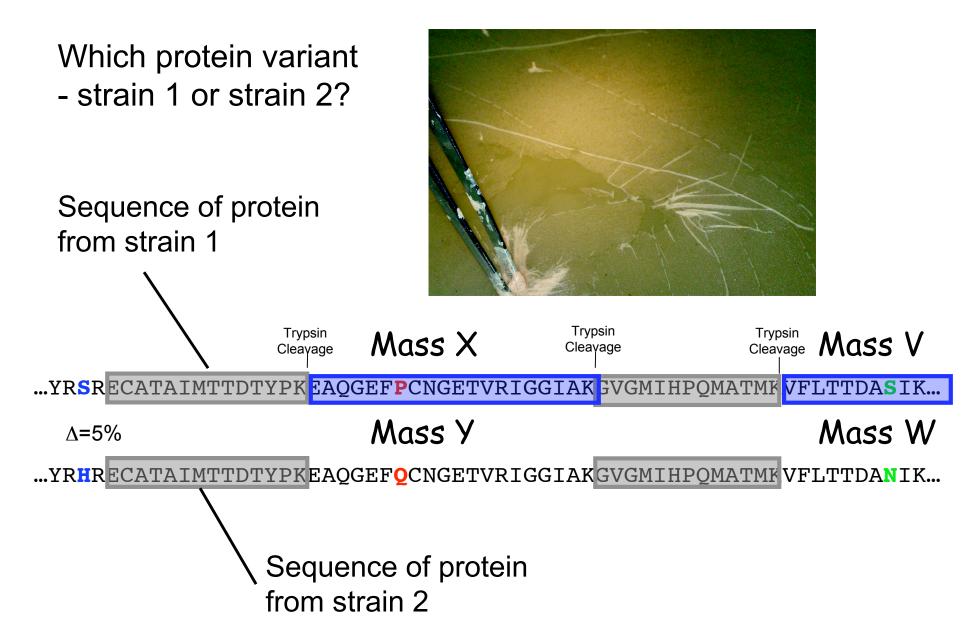








STRAIN-RESOLVED PROTEOMICS



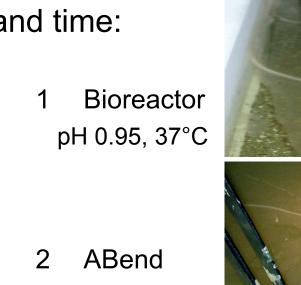
Proteome changes over space and time:

Probable peptidil-prolyl cis-trans isomerase Transcription-repair coupling factor Cytidylyltransferase family protein Phosphoheptose isomerase Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) Protein of unknown function Protein of unknown function Pyruvate synthase alpha subunit (EC 1.2.7.1) Pyruvate synthase beta subunit Putative ferredoxin oxidoreductase gamma subunit Probable ferredoxin oxidoreductase epsilon subunit Protein of unknown function Protein of unknown function Pyruvate synthase alpha chain (EC 1.2.7.1) Putative ferredoxin oxidoreductase beta subunit Ferredoxin oxidoreductase gamma subunit Protein of unknown function Probable ferredoxin oxidoreductase Protein of unknown function Conserved hypothetical protein Glutamyl-tRNA synthetase (EC 6.1.1.17) Protein of unknown function L-aspartate oxidase (EC 1.4.3.16) Putative heat shock protein DnaJ, N-terminal Probable heat shock transcription regulator Chaperone clpB Probable bacterial regulatory protein, Fis family Putative histone-like DNA-binding protein TRNA Val GAC 529025..529099 (+) Putative heavy metal efflux pump, CzcA family Putative secretion protein HlyD Protein of unknown function Putative Xaa-Pro aminopeptidase Probable O-methyltransferase family protein Putative nitrite reductase small subunit Hypothetical Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) Putative glycoside hydrolase, family 57 4-alpha-glucanotransferase (EC 2.4.1.25) Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) Probable aldolase Putative glycoside hydrolase, family 57 Secretion protein HlyD Putative acriflavin resistance protein Putative aminomethyltransferase Putative two component, sigma54 specific, transcriptional regulator, Fis family Probable secretion protein HlyD

ABend pH 1.07, 43°C

3 ABfront pH 0.99, 39°C

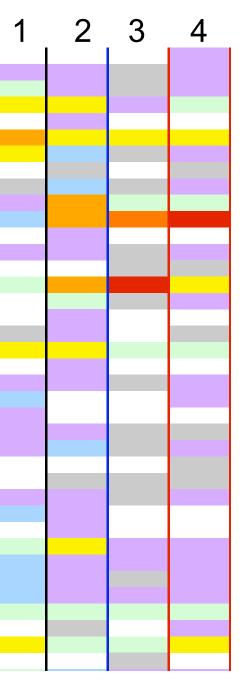
UBA 4 pH 1.28, 38°C





Proteins of unknown function

Protein of unknown function Protein of unknown function



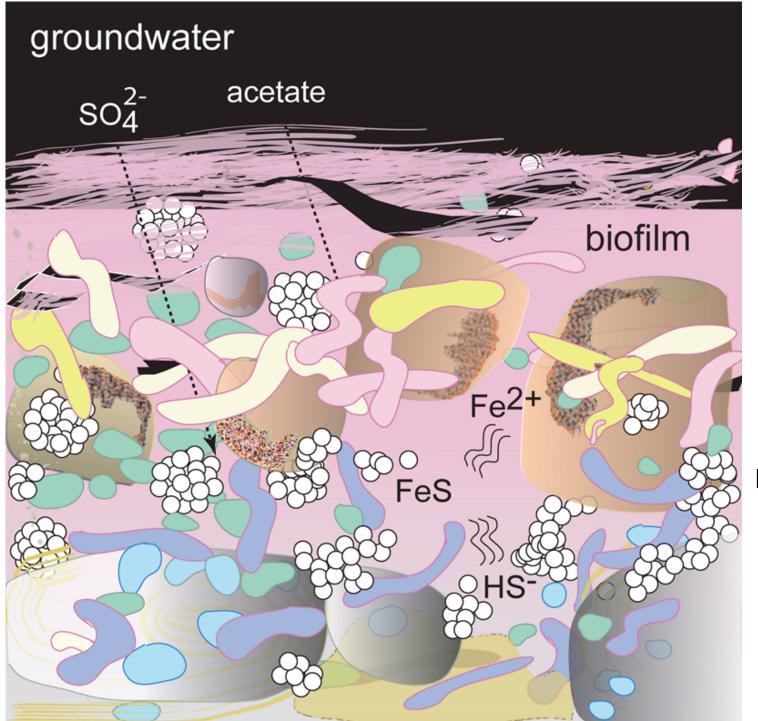
1 Bioreactor pH 0.95, 37°C

2 ABend pH 1.07, 43°C

3 ABfront pH 0.99, 39°C

4 UBA pH 1.28, 38°C





Microbial reduction of ferric ironbearing minerals (Fe²⁺ + S²⁻ = FeS

Microbial sulfate reduction

Old Rifle Aquifer Sediment Columns Studies

Acetate: 3 mM, Sulfate: 9 mM, Flow Rate: 0.3 m/day



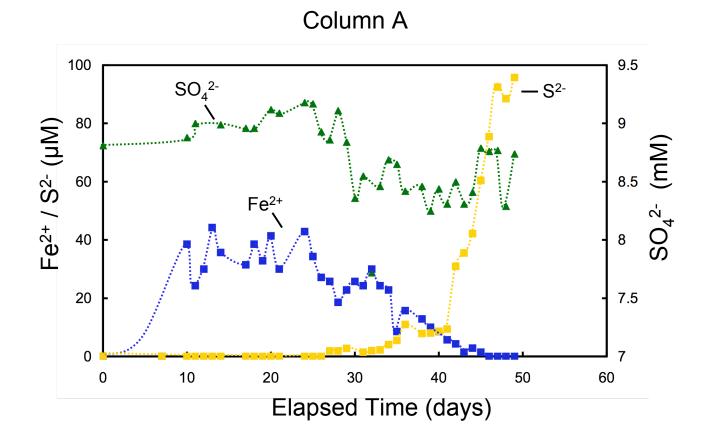
Start

48 days

Mike Wilkins, Ken Williams

Column Studies

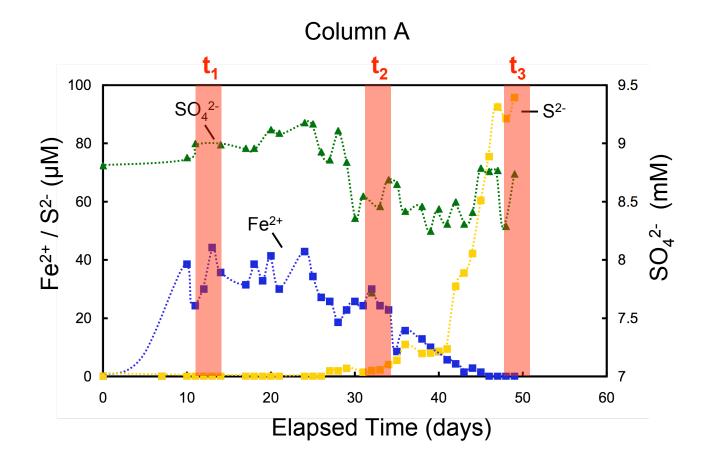
Protein sampling intervals *and locations* determined by pre-determined geochemical - *and geophysical* - indicators



Mike Wilkins, Ken Williams

Column Studies

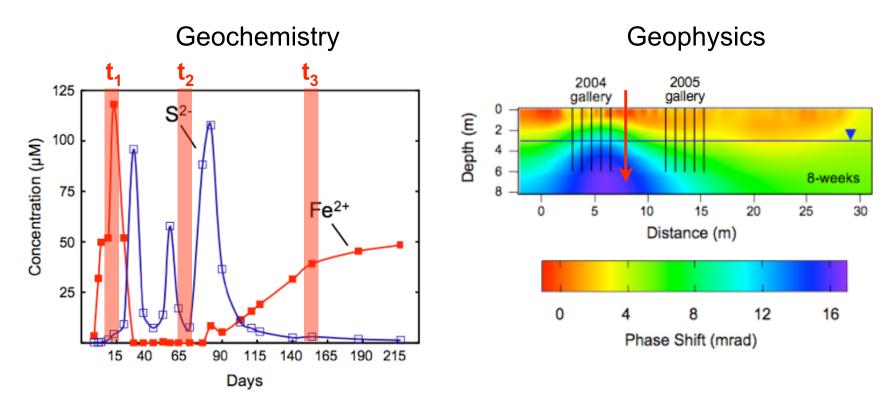
Protein sampling intervals *and locations* determined by pre-determined geochemical - *and geophysical* - indicators



Ken Williams

Field Studies

Protein sampling intervals *and locations* determined by pre-determined geochemical - *and geophysical* - indicators



2006 Old Rifle Field Data

Ken Williams

Rifle sediment proteomics

Proteins to be recovered from batch and column sediments, tubing, filters



M. Wilkins and K. Williams

Batch Incubations

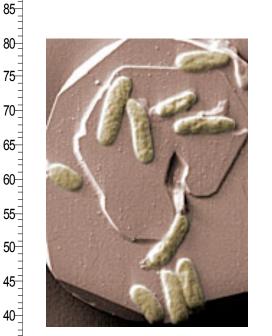
Tubing / Effluent

Protein extraction: protocols for sediment under development

Metagenomic data: from Holmes and Lovley and sequencing to be requested from JGI

Test case: filtered groundwater samples from Rifle, summer 2007

³⁵ Challenges of community proteomics in soil samples



100

90-

35_

30_

25

- Extraction of microbes from soil
- Acquisition and analysis of appropriate genomic sequence
- Interference from humic acids and other organics
- Reabsorption of the proteins to the soils

Quantitative proteomics requires reproducible sample extraction

20				909.9599	
15- 10-					
5					910.4616
90)7.1991 	907.7317	908.6674	909.7623	910.5886 910.9604

U.S. Department of Energy Office of Science

Environmental Remediation Sciences Program

Rifle project protegenomics subgroup:

Mike Wilkins, UC Berkeley Ken Williams, UC Berkeley Greg Dick, UC Berkeley Nathan VerBerkmoes, ORNL Mary Lipton, PNNL Dawn Holmes, U Mass Robert Hettich, ORNL Derek Lovely, U Mass Phil Long, PNNL Jill Banfield, UC Berkeley



Invironmental

Remediation Sciences

AMD proteogenomics research:

Nathan VerBerkmoes, Vincent Denef, Michael Thelen, Robert Hettich, Jill Banfield and collaborators